ANALYZING PRIMARY DATA

Now that you have collected your primary data, it is time to determine what that data means and what you can learn from it. The key when analysing your data is to pull out information that is the most pertinent to your writing: information you can highlight and discuss, and information that will support your claims (if you are making any).

✓ Interviews

Interviews are fairly easy to analyse, as you simply have to go back through the answers you received and decide how to use them within your writing. You can group the answers into categories and create a chart of how those answers may best fit within your paper or article.

If you recorded the interview with a tape or digital recorder, you may want to listen to it and type a transcript of the interview. Since transcription is a tedious process, only use this option if you need to do so.

✓ Surveys

When analysing surveys, you want to get the raw data into form that you can manipulate. If you were using a numerical system or *yes/no* answer system for your survey, you may find it helpful to enter the results into a spreadsheet program such as Microsoft Excel. If the survey was an open-ended question style, see if you can fit your answers into categories of responses.

✓ Observations

Observations are more difficult to analyse because when you are taking notes, you often write down everything that you see. Start by organizing your notes into categories or by some criteria. Once you have everything organized, see if you can make some generalizations about what you have observed.

Over-generalizing Your Results

Your first attempts at primary research will most likely include small groups of people and may not be representative of the population as a whole. It is important to remember not to over-generalize your findings. In other words, do not assume that your findings are necessarily true of every person within the group or every person in a society.

Triangulation of Data

One of the benefits of combining primary research with secondary research is in the area of data triangulation. Data triangulation is when a piece of data, a finding, or a generalization is able to be verified with several different research methods. This helps add to your credibility and makes your findings stronger.

For example, you are studying binge drinking on campus. You find national averages that indicate that 45% of college students binge drink nationwide. You conduct your own research at the Purdue campus. You find that 47% of the individuals you surveyed drink; you also interview a counsellor on campus who reports that approximately 1/3 of the students who he sees suffer from a drinking problem. Thus, your results from an interview with an expert and your own survey support the national averages.

Source: "Conducting Primary Research." *OWL Prudue Online Writing Lab*. Purdue University, 2010. Web. 2 Jan. 2012

NOTES

DATA PREPARATION, INTERPRETATION AND ANALYSIS

Analysing survey data is an important and exciting step in the survey process. It is the time that you may reveal important facts about your respondents, uncover trends that you might not otherwise have known existed, or provide irrefutable facts to support your plans. By doing in-depth data comparisons, you can begin to identify relationships between various data that will help you understand more about your respondents, and guide you towards better decisions.

Three Common Mistakes

Before you dive into analysing your survey results, take a look back at the big picture. What objectives were you trying to accomplish when you created your survey? Did your survey instrument meet those objectives? Is the data you collected the right data? Do you have sufficient data to properly reach a conclusion?

Although data analysis is the wrong time to try and rewrite your survey instrument, it is important to remember the scope of your project and stick to it. Many first time surveyors attempt to read "between the lines" while analysing data. They attempt to answer questions that were not asked by making inferences and assumptions from those that were asked. Doing so amounts to nothing more than guesswork. To avoid this temptation, remember this simple rule:

☞ **If you did not ask, you do not know.**

Inexperience surveyors may attempt to change data to compensate for poor question design. For example, if a question asked a respondent to indicate his total household income using a scale of values, a mean and median cannot be calculated. Many people try to get around this by assigning each response a value representing the range. Even if the adjustment is made consistently across all responses, the resulting calculations will be wrong. Similarly, trying to analyse a multiple-choice

question as if it were a single-select question will often provide erroneous information. In order to avoid this pitfall, remember this simple rule:

☞ **Do not alter data to compensate for bad survey design.**

Inexperienced surveyors may project the findings to an audience that was not either part of the survey population or not adequately represented. For example, if a Human Resource manager conducts a benefits survey and invites all employees to participate, most people would assume that the results represent all employees since everyone had an opportunity to participate. Provided that enough employees participate, the data might be statistically valid, but is it really representative of all employees? The answer is this: it depends. If the survey collected data about employee demographics that could be compared to what is known about the company, then the results do reflect the company as a whole. However, if 80% of the respondents are married and 50% of the total employee base is married, the results of the survey are skewed toward married people. If married people have different benefits needs than single people, using the survey results to make conclusions about the entire employee pool would be less accurate than those conclusions about the married employees or single employees independently. To avoid this temptation, remember this simple rule:

☞ **Do not project your data to people that did not respond.**

The earlier you recognize flaws in your survey design and data collection, the more time you will save during analysis. If your questions do not provide the data you need to meet your survey objectives, you will have to start over. If your questions are vague or ambiguous, you will have to throw them out. If you do not have an adequate number of responses, you will have to get more.

Survey Analysis

Analysing any survey, web or traditional, consists of a number of interrelated processes that are intended to summarize, arrange, and transform data into information. If your survey objective was simply to collect data for your database or data warehouse, you do not have to do any analysis of the data. On the other hand, if your objective was to understand the characteristics of typical individuals, then you must transform you raw results into information that will enable you to paint a clear picture of those individuals.

Assuming you need to analyse the data collected from your survey, the process begins with a quick review of the results, followed by editing, analysis, and reporting. To ensure you have accurate data before investing significant time in analysis, it is important that you do not begin analysing results until you have completed the review and editing process.

❖ Quick Review

Read all your results. Although this seems like an obvious thing to do, many surveyors think that they can skip this step and dive right into data analysis. A quick review can tell you a lot about your project, including any flaws in questionnaire design or response population, before you spend hours of time in analysing the data.

During the quick review, you should look at every question and see if the results "make sense". This "gut feel" check of the data will often uncover any issues with your survey project. Most surveyors already have an idea of how they expect their data to look. A quick review of the data can help you quickly understand and tell you if the people that respond are the right people. For example, if you were conducting a survey of all the employees in a company and you knew that 10% were in the marketing department, 20% in sales, 45% in manufacturing, 5% in management, and 5% finance, and 15% research and development, you could reasonably expect your responses to be similarly distributed. If your quick review disclosed 80% of your respondents were from the sales department, you know that your survey did not adequately capture a representative sample of all departments within the company.

The quick review can also highlight any problems with the survey instrument. Are most respondents answering all questions? If not, your questionnaire could be flawed in such a way that a person cannot complete the survey. A low response rate could mean your survey invitation was not compelling enough to encourage participation, or your timing was off and a follow-up reminder is needed.

Lastly, the quick review of the survey can show you what areas to focus on for detailed analysis. As stated earlier, most surveyors already know what they expect to get, so your quick review can show you the unexpected.

❖ Editing and Cleaning

Editing and cleaning data is an important step in the survey process. Special care must be taken when editing survey data so that you do not alter or throw out responses in such a way as to bias your results. Although you can begin editing and cleaning your data as soon as results are received, caution should be used since any edits can be lost if the database is rebuilt. To be safe, wait until all data is received before you begin the editing and cleaning process.

To start, find and delete incomplete and duplicate responses. A response should be discarded if the respondent did not complete enough of the survey to be meaningful. For example, if your survey was intended to determine future buying intentions across various demographic groups and the respondent did not answer any of the demographic questions, you should delete the

response. On the other hand, if the respondent answered all the demographic questions but omitted his/her name or email address, then you should keep the response.

Data cleaning of web surveys usually involves categorizing answers to open-ended questions and multiple-choice questions that include an “other, please specify” response. Because of their nature, open-ended text response questions can provide significant value but they are nearly impossible to process without some form of summarization or tabulation. One of the easiest ways to summarize these questions is to build a list of themes and select the themes that apply as you read each response.

A common problem in any survey that needs attention during the editing and cleaning process is when a respondent answers an “other, please specify” question by selecting “other” and then writing in an answer that was one of the listed response options. Without cleaning these answers, the “other” response will be overstated and the correct response will be understated. For example, a demographics question that asks for the respondent’s role within the organization may have a response like “faculty, teacher, or student” and a respondent selects “other” and types “professor”, you would want to clean the response by switching the other choice to the one for “faculty, teacher, or student”.

Once the data preparation is complete, it is time to start analysing the data and turning it into actionable information.

❖ Detailed Analysis

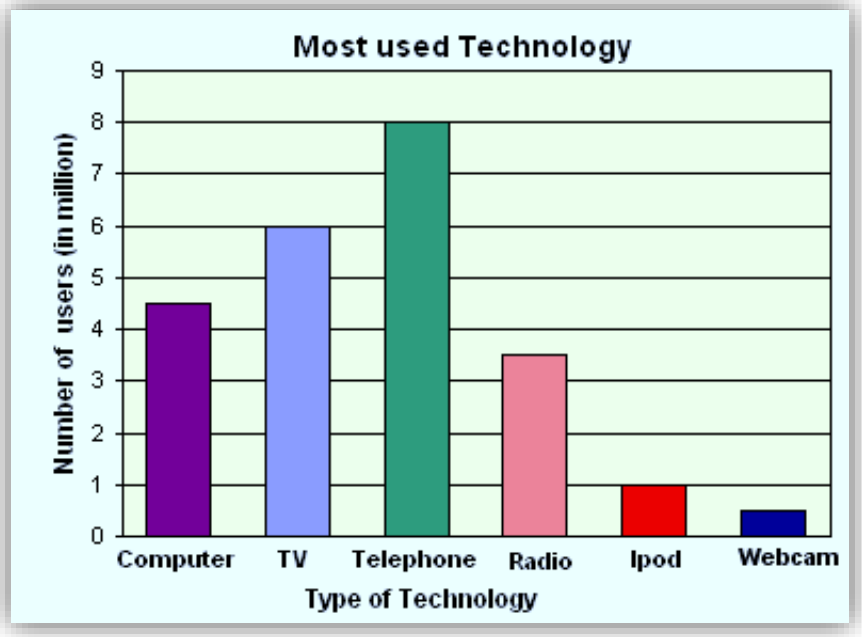
Analysis is the most important aspect of your survey research project. At this point, you have collected a set of data that must now be turned into actionable information. The process of analysis can lead to a variety of alternative courses of action. Mistakes during analysis can lead to costly decisions down the road, so extreme caution and careful review must be followed throughout the process. Carelessness during analysis can lead to disaster. What you do during analysis will ultimately determine if your survey project is a successful one or not.

Depending on what type of information you are trying to know about your audience, you will have to decide what analysis makes sense. For most surveyors, a basic analysis using charts, cross tabulations, and filters is sufficient. On the other hand, more sophisticated users may wish to do a more complex statistical analysis using high powered analytical tools such as SPSS, Excel, or any number of number crunching applications. For our purposes in this article, we will focus on basic analysis techniques.

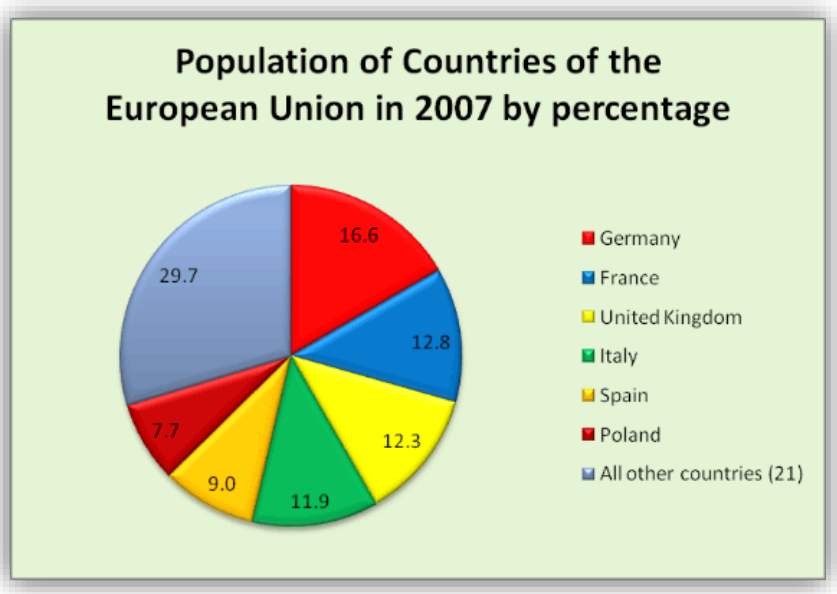
Graphical Analysis

Graphical analysis simply means displaying the data in a variety of visual formats that make it easy to see patterns and identify differences among the results set. There are many different graphing options available to display data, the most common are *Bar*, *Pie*, *Line Charts* and *Histograms*.

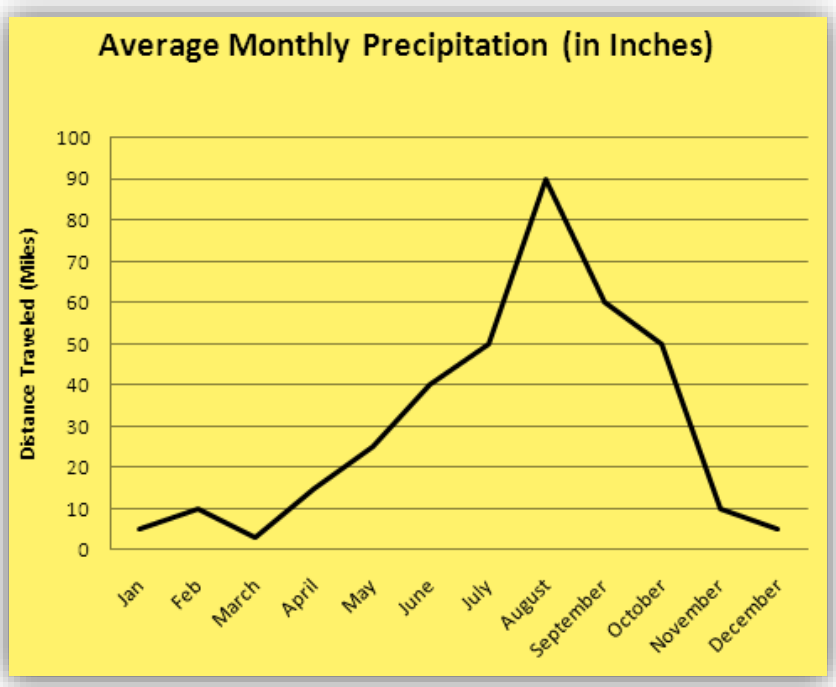
Bar charts use solid bars on an X and Y-axis that extend to meet a specific data value indicated on the chart and can be shown either vertically or horizontally. These charts are flexible and are most commonly used to display data from multiple-select, rank order, single-select matrix and numerical questions. Each response option is shown as an independent bar on the chart, and the length of the bar represents the frequency the response was chosen relative to all choices.



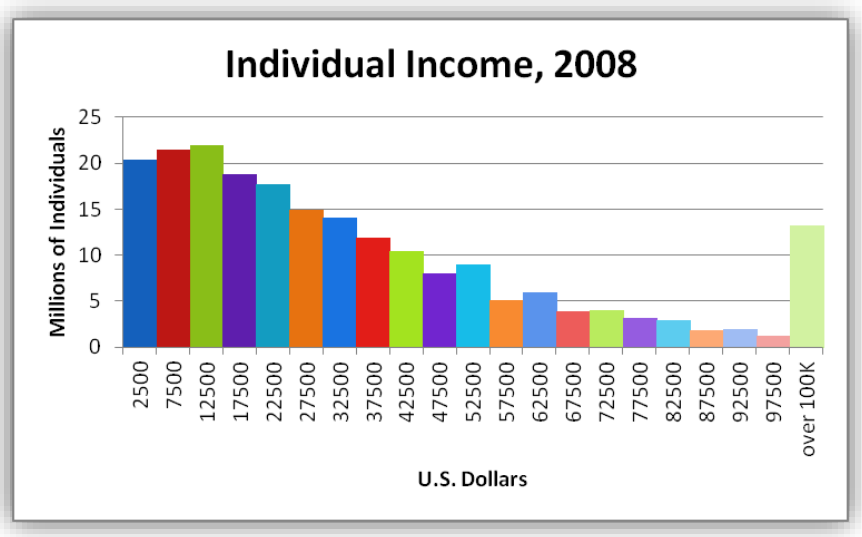
Pie charts, or circle graphs, have colourful "slices" representing segments of your data. These charts measure values as compared to a "whole", and the total percentages of the segments always add up to 100%. Pie charts are most useful with single-select questions because each response is represented visually as a portion of the entire pie. It is easy to interpret which answer received the most responses in a pie chart by selecting the largest portion of the pie. When comparing two sets of data using a pie chart, it is important to make sure the colours used for each response option remain consistent in each chart. Pie charts are not appropriate for multiple-select questions because each respondent can choose more than one option, and the sum of the option percentages will exceed 100%.



A **line graph** displays the relationship between two types of information, such as number of school personnel trained by year. They are useful in illustrating trends over time.



A **histogram** has connected bars that display the frequency or proportion of cases that fall within defined intervals or columns. The bars on the histogram can be of varying width and typically display continuous data.



There are other graphing options such as area charts and scatter graphs, which are useful when displaying the same data over a period of time. However these formats are not as easy to interpret for casual users, so they should be used sparingly.

Frequency tables are another form of basic analysis. These tables show the possible responses, the total number of respondents for each part, and the percentages of respondents who selected each answer. Frequency tables are useful when a large number of response options are available, or the differences between the percentages of each option are small. In most cases, pie or bar charts are easier to work with than frequency tables.

Response	Count	Percent
Market Analysis	76	13.7%
Quantitative Analysis	150	27.0%
Strategic Planning	56	10.1%
Product Planning	33	5.9%
Promotional Communication	243	43.8%
Creating sales tools	152	27.4%
Providing channel support	157	28.3%

Cross tabulations, or cross tabs, are a good way to compare two subgroups of information. Cross tabs allow you to compare data from two questions to determine if there is a relationship between them. Like frequency tables, cross tabs appear as a table of data showing answers to one question as a series of rows and answers to another question as a series of columns.

Base Question	Female	Male
Product Manager	57.2%	53.4%
Director	12.6%	14.2%
Product Marketing Manager	24.7%	23.1%
Program Manager	2.8%	1.5%
Technical Product Manager	2.8%	7.7%
Total Counts	215	337

Cross tabs are used most frequently to look at answers to a question among various demographic groups. The intersections of the various columns and rows, commonly called cells, are the percentages of people who answered each of the responses. In the example above, females and males had relatively similar distribution among various job titles, with the exception of the tile of "Technical Product Manager", where 2.5 times as many males had the title as compared to females. For analysis purposes, cross tabs are a great way to do comparisons.

Filtering is the most under-utilized tool used in analysis. Filters allow you select specific subsets of data to view. Unlike a cross tab, that compares two questions, a filter will allow you to examine all questions for a particular subset of the responses. By viewing only the data from the people who responded negatively, look at how they answered other questions. Find patterns or trends that help define why a person answered the way they did. You can even filter on multiple questions and criteria to do a more detailed search if necessary. For example, if you wanted to know the buying intentions of men, over the age of 40, with income of about $50,000, you would set a filter that would remove all those respondents that do not meet your criteria from the results set, thus enabling you to concentrate on the target population.

By applying filters to the date survey responses were received, you can see how the answers change from one time frame to the next. For instance, by continually running a customer satisfaction survey, you can assess changes in customer attitudes over time by filtering on the date the survey was received. You can also use a filter on date received to assess the impact of sales incentive programs or new product offerings by comparing survey responses before and after the change.

Filters do not permanently remove the responses of those people that do not match the specified criteria; they simply eliminate them from the current view of the data, making it much easier to perform analysis. By looking at the same question with different filters applied, differences between the various respondents represented by the filter can be quickly seen. Because filters remain in effect until cleared, don't forget to clear them before attempting to analyze your survey responses as a whole, otherwise your observations will be inaccurate, and your recommendations flawed.

Simple Regression Analysis – Determining what factors have led to a particular outcome is called *regression analysis*. The regression means that you are working backwards from the result to find out why a person answered the way that he/she did. This can be based on how he/she answered other questions as well.

For example, you might believe that website visitors who had trouble navigating within your website are likely not to return again. If 30% of the respondents said they had trouble navigating through the website and 40% said they would not return, you could look at only those that would not return to determine if poor navigation might be the case. After filtering to only those who would not return, you may discover that 30% or less said they had trouble navigating; this is clearly not the main "reason" for visitors not returning. By filtering out those that would return, we expect the percentage to increase dramatically. If it does, we still cannot conclude that navigation is "the" reason; only that it might contribute to the respondents not returning. In order to know if it is "the" reason, we would need to ask a direct question.

❖ Reporting

After analyzing your survey data, it is time to create a report of your findings. The complexity and detail needed to support your conclusions, along with your intended audience, will dictate the format of your report. CEO's require a different level of detail than line managers, so for maximum results consider who is going to receive your report and tailor it to meet their unique needs.

Visual reports, such as an HTML document or Microsoft PowerPoint presentation, are best suited for simple findings. These graphical reports are best when they are light on text and heavy on graphs and charts. They are reviewed quickly rather than studied at length, and most conclusions are obvious, so detailed explanations are seldom required. For more complex topics, a detailed report created in Microsoft Word or Adobe Acrobat is often required. Reports created using Word often include much more detailed information, report findings that require significant explanation, are extremely text heavy, and are often studied at great length and in significant detail.

No matter which type of report you use, always remember that information can be more powerfully displayed in a graphic format verses a text or tabular representation. Often, trends and patterns are more obvious and recommendations more effective when presented visually. Ideally, when making comparisons with one or more groups of respondents, it is best to show a chart of each group's responses side-by-side. This side-by-side comparison allows your audience to quickly see the differences you are highlighting and will lead to more support for your conclusions.

At the beginning of your report, you should review your survey objective and sampling method. This will help your audience understand what the survey was about, and enable you to avoid many questions that are outside of your original objectives. Your report should have a description of your sampling method, including who was invited to participate, over what time frame results were collected, and any issues that might exist relative to your respondent pool. Next, you should include your analysis and conclusions in adequate detail to meet the needs of your audience. Include a table or graph for each area of interest and explain why it is noteworthy. After your analysis section, you should make recommendations that relate back to your survey objectives. Recommendations can be as simple as conduct further studies to a major shift in company direction. In either case, your recommendation must be within the scope of your survey objective and supported by the data collected. Finally, you can include a copy of your survey questions and a summary of all the data collected as an appendix to your report.

❖ Conclusion

Survey analysis is not as easy as downloading results and printing a chart or report, yet it is not so complex that it requires a PhD. In this article we have learned that good analysis begins with good questions, representative participation, and careful interpretation of the data, in order to produce actionable results. Techniques such as charting, filtering, cross tabulation, and regression analysis all help you spot trends and patterns within your data while helping you meet your survey objective.

Guidelines for Formatting Graphs and Charts

- **Keep it simple and avoid flashy special effects.** Present only essential information**.** Avoid using gratuitous options in graphical software programs, such as three-dimensional bars, that confuse the reader. If the graph or chart is too complex, it will not clearly communicate the important points.

- **Title your graph or chart clearly to convey the purpose.** The title provides the reader with the overall message you are conveying.

- **Specify the units of measurement on the x-and y-axis.** Years, number of participants trained, and type of school personnel are examples of labels for units of measurement.

- **Label each part of the chart or graph.** You may need a legend if there is too much information to label each part of the chart or graph. (See the line graph). Use different colors or variations in patterns to help the reader distinguish categories and understand your graph or chart.

Source: Minter E, Michaud M. *Using Graphics to Report Evaluation Results*. University of Wisconsin Cooperative Extension. 2003. Available at http://learningstore.uwex.edu/pdf/G3658-13.PDF

NOTES